2014 JMD CME Program

The Journal of Molecular Diagnostics, Vol. 16, No. 2, March 2014

ELSEVIER

the Journal of Molecular Diagnostics

jmd.amjpathol.org

Molecular Diagnosis of Autosomal Dominant Polycystic Kidney Disease Using Next-Generation Sequencing

Adrian Y. Tan,* Alber Michaeel,* Genyan Liu,* Olivier Elemento,† Jon Blumenfeld,‡§ Stephanie Donahue,§ Tom Parker,§ Daniel Levine,§ and Hanna Rennert*

From the Departments of Pathology and Laboratory Medicine, Physiology and Biophysics,† and Medicine,‡ Weill Cornell Medical College, New York; and The Rogosin Institute,§ New York, New York*

CME Accreditation Statement: This activity ("JMD 2014 CME Program in Molecular Diagnostics") has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint sponsorship of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("JMD 2014 CME Program in Molecular Diagnostics") for a maximum of 48 *AMA PRA Category 1 Credit(s)*™. Physicians should only claim credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Accepted for publication October 31, 2013.

Address correspondence to Hanna Rennert, Ph.D., Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, 1300 York St, F701, New York, NY 10021. E-mail: har2006@med.cornell.edu.

Autosomal dominant polycystic kidney disease (ADPKD) is caused by mutations in *PKD1* and *PKD2*. However, genetic analysis is complicated by six *PKD1* pseudogenes, large gene sizes, and allelic heterogeneity. We developed a new clinical assay for PKD gene analysis using paired-end next-generation sequencing (NGS) by multiplexing individually bar-coded long-range PCR libraries and analyzing them in one Illumina MiSeq flow cell. The data analysis pipeline has been optimized and automated with Unix shell scripts to accommodate variant calls. This approach was validated using a cohort of 25 patients with ADPKD previously analyzed by Sanger sequencing. A total of 250 genetic variants were identified by NGS, spanning the entire exonic and adjacent intronic regions of *PKD1* and *PKD2*, including all 16 pathogenic mutations. In addition, we identified three novel mutations in a mutation-negative cohort of 24 patients with ADPKD previously analyzed by Sanger sequencing. This NGS method achieved sensitivity of 99.2% (95% CI, 96.8%—99.9%) and specificity of 99.9% (95% CI, 99.7%—100.0%), with cost and turnaround time reduced by as much as 70%. Prospective NGS analysis of 25 patients with ADPKD demonstrated a detection rate comparable with Sanger standards. In conclusion, the NGS method was superior to Sanger sequencing for detecting PKD gene mutations, achieving high sensitivity and improved gene coverage. These characteristics suggest that NGS would be an appropriate new standard for clinical genetic testing of ADPKD. *(J Mol Diagn 2014, 16: 216—228; http://dx.doi.org/10.1016/j.jmoldx.2013.10.005)*

Autosomal dominant polycystic kidney disease (ADPKD) affects 1 in 400 to 1 in 1000 live births worldwide.[1] It is the most common inherited kidney disease, accounting for approximately 5% of the end-stage renal disease population.[2] ADPKD is initiated by gene mutations in renal tubular epithelial cells, which seem to be more sensitive to haploinsufficiency, resulting in increased proliferation and cyst formation.[3] The consequent increase in the number and size of kidney cysts causes progressive chronic kidney disease.[4] ADPKD is mainly caused by mutations in two large genes, *PKD1* and *PKD2*, accounting for 75% to 85% and 15% to 25% of cases, respectively, in clinically well-characterized populations. *PKD1* spans 46 exons and encodes polycystin-1 with 4303 amino acids.[5] *PKD2* spans 15 exons, encoding polycystin-2, which consists of 968 amino acids.[6] Chromosome 16 includes six homologous genes (ie, pseudogenes) that share 97.7% sequence identity with the *PKD1* gene exons 1 to 33.[7,8]

Supported by The Rockefeller University Center for Clinical Translational Science (J.B.), National Center for Research Resources/NIH grant UL1 RR024143-01 (J.B.), the NIH Roadmap for Medical Research (J.B.), and by the Starr Foundation (J.B.).

Disclosures: None declared.

http://dx.doi.org/10.1016/j.jmoldx.2013.10.005

The clinical diagnosis of ADPKD is established by family history and renal imaging modalities, such as ultrasonography, computed tomography, and magnetic resonance imaging.[9] However, these diagnostic test results are often ambiguous, particularly in young individuals. Consequently, genetic testing plays an increasingly important role in the diagnosis and management of patients with ADPKD.[10] Moreover, with the development of potentially effective pharmacologic treatments for ADPKD,[11] the need for accurate diagnostic genetic tests has become more compelling.

The key step in ADPKD genotyping procedures is amplification of the *PKD1* gene region while excluding the pseudogenes. This was traditionally achieved by using long-range PCR (LR-PCR) with primers located to the rare mismatch sites that distinguish *PKD1* and the pseudogenes, followed by nested PCR of the individual exons, whereas the single-copy regions of *PKD1* and *PKD2* were directly amplified from genomic DNA. Amplicons were then directly analyzed by Sanger sequencing or by sequencing coupled with a mutation screening step to lower the testing cost.[12,13] However, the genetic analysis of ADPKD is challenging, especially owing to the large size, complex genomic structure, and allelic heterogeneity of *PKD1* and *PKD2*.[14] Next-generation sequencing (NGS) technology has revolutionized the field of human genetics and molecular diagnostics.[15,16] Recently, Rossetti et al[17] reported a mutation screening strategy for analyzing PKD genes using NGS by pooling LR-PCR amplicons and multiplexing bar-coded libraries. This approach was designed to have a high throughput and has been successful for screening mutations in large cohorts. However, the method had low sensitivity and slow turnaround time, mainly because of the sample pooling strategy used.

Herein, we present a new NGS-based genotyping approach for patients with ADPKD that is better tailored to the standard clinical diagnostic setting, where rapid turnaround time and high sensitivity could be achieved by individually bar coding each patient in the run. The diagnostic performance of the new assay was evaluated using a panel of DNA samples previously analyzed by Sanger sequencing.[18] The testing strategy, workflow, data analysis pipeline, costs, and other related issues are also discussed.

Materials and Methods

Study Patients

Study patients were participants in The Rogosin Institute Polycystic Kidney Disease Data Repository (*http://www.clinicaltrials.gov*, Identifier NCT00792155). This is a single-center, prospective, longitudinal study of genotype and phenotype characteristics of individuals with ADPKD. Study samples were randomly selected for analysis. All the participants underwent PKD genotyping by the Weill Cornell Medical College Molecular Pathology Research Laboratory (New York, NY) using direct sequencing or SURVEYOR nuclease−WAVE screening (Transgenomic Inc., Omaha, NE). In addition, we prospectively analyzed a new cohort of patients with ADPKD that has not been previously genotyped. The study was approved by the Institutional Review Board Committees at Weill Cornell Medical College and The Rockefeller University (New York, NY). All the participants provided written informed consent.

Long-Range PCR

Genomic DNA was extracted from peripheral blood lymphocytes using a Gentra Puregene blood kit (Qiagen Inc., Valencia, CA). The entire coding region, the exon-intron boundaries, and most of the 5′ and 3′ untranslated regions of *PKD1* and *PKD2* were amplified in a total of 10 (five reactions per gene) distinct PCR reactions using primers anchored either in the rare mismatched region with the human homologs or in the single-copy region of *PKD1*. The LR-PCR primers were designed using Primer3 software version 4.0.0 (*http://bioinfo.ut.ee/primer3*, last accessed October 25, 2013) (Table 1).[19] The LR-PCR primers were modified at the 5′ end with NH2 to prevent overrepresentation of sequences at the amplicon ends in the ligation step and to increase sequence coverage uniformity.[20] LR-PCR was performed using either the GeneAmp high fidelity PCR system (Applied Biosystems, Foster City, CA) or the PrimeSTAR GXL DNA polymerase (Takara Bio Inc., Shiga, Japan). PCR amplification conditions for the various LR-PCR fragments are described in Supplemental Table S1. After purification with Agencourt AMPure XP beads (Beckman Coulter, Beverly, MA), the LR-PCR fragments from each patient were quantified using PicoGreen (Quant-iT; Invitrogen, Carlsbad, CA) and were pooled together at equal molar ratio.

Library Preparation and Indexing

For each patient, 4 μg of LR-PCR products were pooled together in a total reaction volume of 210 μL of Tris-EDTA buffer and were fragmented to approximately 300 bp using adaptive focused acoustics (Covaris S2; LGC Ltd., Teddington, UK) with the following settings: duty cycle, 20%; intensity, 5; and cycles per burst, 200. After shearing, the fragments underwent end repair using the NEBNext end repair module (New England BioLabs Inc., Ipswich, MA) by adding 3′ dA overhangs to the blunt-ended DNA. After purification on AMPure XP beads, unique indexed adaptors were ligated to each patient pool using the NEBNext quick ligation module (New England BioLabs Inc.). Twenty-five different adapters with 5-nt bar codes were used for indexing and library preparation. The bar codes located in the 3′ end of each adapter were designed using a published Python script (create_index_sequences.py)[21] with a minimum edit distance, or mutation tolerance, of three. Each adapter sequence began with the paired-read oligonucleotide sequences as specified by Illumina Inc. (San Diego, CA): 5′-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3′ and 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′. Annealing of the indexed adaptors was performed at 95°C for 2 minutes, followed by a cooldown to room temperature at a rate of 0.1°C per second using a thermal cycler (Biometra GmbH,